



DataShare REFLIB Documentation

DataShare Programming and Informatics Team

August 8, 2023

TABLE OF CONTENTS

REVISION HISTORY	3
I. PURPOSE.....	4
II. UNIFIED MEDICAL LANGUAGE SYSTEM (UMLS).....	4
A. BERENSON-EGGERS TYPE OF SERVICE (BETOS).....	5
B. CODE ON DENTAL PROCEDURES AND NOMENCLATURE (CDT)	5
C. CURRENT PROCEDURAL TERMINOLOGY (CPT) / HEALTHCARE COMMON PROCEDURE CODING SYSTEM (HCPCS).....	6
D. INTERNATIONAL CLASSIFICATION OF DISEASES, 9TH REVISION, CLINICAL MODIFICATION (ICD-9-CM) - DIAGNOSIS AND PROCEDURE	7
E. INTERNATIONAL CLASSIFICATION OF DISEASES, 10TH REVISION, CLINICAL MODIFICATION (ICD-10-CM)	9
F. INTERNATIONAL CLASSIFICATION OF DISEASES, 10TH REVISION, PROCEDURE CODING SYSTEM (ICD-10-PCS) ..	10
III. OTHER DATA RESOURCES	11
A. RESTRUCTURED BETOS CLASSIFICATION SYSTEM (RBCS).....	11
B. ICD-10-CM AND ICD-10PCS GENERAL EQUIVALENCE MAPPINGS (GEMS).....	12
C. CLINICAL CLASSIFICATION SOFTWARE - REFINED (CCSR) - DIAGNOSIS AND PROCEDURE VERSIONS	13
D. CLINICAL CLASSIFICATION SOFTWARE – SERVICES AND PROCEDURES VERSION (CPT/HCPCS).....	14
E. PROVIDER OF SERVICE (POS)	15
F. MULTIPLE SEVERITY - DIAGNOSES RELATED GROUPS (MSDRG)	16
G. NATIONAL PLAN AND PROVIDER ENUMERATION SYSTEM (NPPES)	17
H. NATIONAL DRUG CODE DIRECTORY (NDC).....	19
I. AREA HEALTH RESOURCE FILES (ARHF).....	21
J. CROSSWALK FOR ZIP CODE TO HOSPITAL SERVICE AREA (HSA) AND HOSPITAL REFERRAL REGION (HRR)	22
K. AREA DEPRIVATION INDEX (ADI) - 9-DIGIT ZIP CODE LEVEL.....	23

REVISION HISTORY

Version	Date	Description	Revised By
1.0.0	1/12/2023	Initial version for Fall 2022 refresh	Duke University Department of Population Health Sciences
1.0.1	4/23/2023	Added CCS CPT procedure table	Duke University Department of Population Health Sciences
1.0.2	8/3/2023	Updated BETOS, CDT, and CPT_HCPCS tables	Duke University Department of Population Health Sciences

I. PURPOSE

DataShare curates a library of reference tables for coding systems and standards that are commonly used in biomedical data and research. These tables are updated periodically (typically, in the fall) and are made available to all DataShare users as part of the DataShare infrastructure. The data are housed in the REFLIB schema on the DataShare's Oracle server within PACE. Users with access can view the data using DBVisualizer. To pull/merge these data into a project's analysis files, users can create a SAS libname connection to the Oracle database. Users must have an account on the DataShare Oracle server to access the data. If you have questions about access, please contact pophealthdatashare@dm.duke.edu.

This document describes the data resources that are included in the REFLIB schema. In section II, we describe the reference tables that we extract from the NIH National Library of Medicine's Unified Medical Language System (UMLS). In section III, we describe reference tables that we retrieve from other sources.

II. UNIFIED MEDICAL LANGUAGE SYSTEM (UMLS)

The NIH National Library of Medicine actively maintains a system called the Unified Medical Language System (UMLS) that includes many biomedical standards. In the UMLS, the term "vocabulary" is used to describe a set of classification or coding standards. A large set of vocabularies are combined into a relational database structure called the "Metathesaurus". New versions of the Metathesaurus are release twice a year (spring and fall), although many of the vocabularies are only updated once per year. More information about the UMLS and the Metathesaurus can be found at: [UMLS - Metathesaurus \(nih.gov\)](http://umls.nlm.nih.gov)

The specific UMLS Metathesaurus vocabularies that DataShare extracts for REFLIB are described below.

A. BERENSON-EGGERS TYPE OF SERVICE (BETOS)

Berenson-Eggers Type of Service (BETOS) is a classification standard for procedures and services that aggregates procedure codes from the Health Care Financing Administration Common Procedure Code System (HCPCS; see below for description) into clinically meaningful groups.

Note: The BETOS coding system is being deprecated and replaced with the Restructured BETOS Classification System (RBCS; see below for description).

Table Name	BETOS
Source	UMLS - Metathesaurus (nih.gov) CPT and HCPCS source files
Version currently in REFLIB	May 2023 (UMLS2023AA)
Release Frequency	Annual (Spring release)
Release-specific or cumulative?	Release-specific
Other notes or links	Line Berenson-Eggers Type of Service (BETOS) Code ResDAC

Variable name	Variable description
CODE	HCPCS code
BETOS	BETOS category code
BETOS_DESCRIP	Text description of BETOS category

B. CODE ON DENTAL PROCEDURES AND NOMENCLATURE (CDT)

The Code on Dental Procedures and Nomenclature (formerly known as Current Dental Terminology) system includes procedure codes for dental treatment and services.

Table Name	CDT
Source	UMLS - Metathesaurus (nih.gov)
Version currently in REFLIB	May 2023 (UMLS2023AA)
Release Frequency	Annual (Spring release)
Release-specific or cumulative?	Cumulative - DataShare combines multiple years of CDT code lists into one cumulative table in Oracle so that historical information and codes are retained.
Other notes or links	UMLS Metathesaurus - CDT (CDT) - Synopsis (nih.gov)

Variable name	Variable description
CODE	CDT/HCPCS code
DESCRIP	Text description of service/procedure
DESCRIP_ALT	Abbreviated text description of service/procedure
EVER_ACTIVE	Was the code ever active during the period for which we have UMLS data? (Created by DataShare)
FIRST_RELEASE	UMLS version/release in which the code was first introduced (Created by DataShare)

C. CURRENT PROCEDURAL TERMINOLOGY (CPT) / HEALTHCARE COMMON PROCEDURE CODING SYSTEM (HCPCS)

The Current Procedural Terminology (CPT) system is published by the American Medical Association and includes billing codes for medical treatment and services.

The Healthcare Common Procedure Coding System (HCPCS) coding standard is maintained by CMS. HCPCS leverages the existing CPT system and extends it with additional procedure/service billing codes that are specific to CMS. HCPCS has two levels; Level I includes the full CPT code list, while Level II has the CMS-specific procedure/service codes. In the UMLS Metathesaurus, these two levels are stored in separate vocabularies. However, we merge them together because these values often appear in the same data field in administrative health data.

Table Name	CPT_HCPCS
Source	UMLS - Metathesaurus (nih.gov) CPT, CPT in HCPCS, and HCPCS source files
Version currently in REFLIB	May 2023 (UMLS2023AA)
Release Frequency	Annual (Spring release)
Release-specific or cumulative?	Cumulative - DataShare combines multiple years of CPT/HCPCS code lists into one cumulative table in Oracle so that historical information and codes are retained.
Other notes or links	UMLS Metathesaurus - CPT (CPT - Current Procedural Terminology) - Synopsis (nih.gov) UMLS Metathesaurus - HCPCS (HCPCS - Healthcare Common Procedure Coding System) - Synopsis (nih.gov) CPT® (Current Procedural Terminology) AMA (ama-assn.org)

Variable name	Variable description
CODE	CPT/HCPCS code Note: HCPCS Level 1 (CPT codes) are 5-digit numeric; HCPCS Level II are 5-character codes beginning with a letter and followed by 4 numeric digits.
DESCRIP	Text description of service/procedure
DESCRIP_ALT	Abbreviated text description of service/procedure
EVER_ACTIVE	Was the code ever active during the period for which we have UMLS data? (Created by DataShare)
FIRST_RELEASE	UMLS version/release in which the code was first introduced (Created by DataShare)

D. INTERNATIONAL CLASSIFICATION OF DISEASES, 9TH REVISION, CLINICAL MODIFICATION (ICD-9-CM) - DIAGNOSIS AND PROCEDURE

The International Classification of Diseases, Ninth Revision (ICD-9) system is published by the World Health Organization and includes diagnosis codes for diseases and related health problems. CMS and the National Center for Health Statistics maintain the “Clinical Modification” ICD-9-CM version.

ICD-9-CM codes are structured in a hierarchy, with additional digits after the decimal point signifying more specific codes. For example:

- 403 = Hypertensive chronic kidney disease
- 403.0 = Malignant hypertensive renal disease
- 403.00 = Hypertensive chronic kidney disease, malignant, with chronic kidney disease stage I through stage IV, or unspecified
- 403.01 = Hypertensive chronic kidney disease, malignant, with chronic kidney disease stage V or end stage renal disease

However, not all level of the hierarchy are considered official “billable” codes for CMS claims processing and other administrative systems. In the example above, “403” and “403.0” are not considered billable codes. Note, however, that the “billable” level of hierarchy is different for different code trees. In some cases, just the first 3 or 4 digits may be required (e.g. 061 Dengue)

Caution: Avoid using Microsoft Excel to handle any ICD-9-CM codes because Excel does not handle trailing zeroes correctly. For example, Excel will consider “403”, “403.0”, and “403.00” to be equivalent and just convert all to “403”.

Note: ICD-9-CM was retired by CMS and most U.S. health entities in October 2015. Any claims from October 1, 2015 to present should use ICD-10-CM and ICD-10-PCS codes.

Table Name	ICD9CM_DX and ICD9CM_PX
Source	UMLS - Metathesaurus (nih.gov)
Version currently in REFLIB	UMLS2015AA
Release Frequency	Retired (previously, Annual)
Release-specific or cumulative?	Cumulative - DataShare combines multiple years of ICD-9-CM diagnosis code lists into one cumulative table in Oracle so that historical information and codes are retained.
Other notes or links	UMLS Metathesaurus - ICD9CM (International Classification of Diseases, Ninth Revision, Clinical Modification) - Synopsis (nih.gov)

Variable name	Variable description
CODE	ICD-9-CM diagnosis code (in _DX table) or procedure code (in _PX table) Note: In this table, the decimal point is included in the ICD diagnosis code. However, other administrative billing databases may not include the decimal, so users may need to remove the decimals before attempting to merge/join with this table.
DESCRIP	DX table: Text description of diagnosis for disease or health condition PX table: Text description of procedure
DESCRIP_ALT	DX table: Alternate text description of diagnosis for disease or health condition PX table: Alternate text description of procedure
VALID	Code is valid for billing. (0=No, 1=Yes)

FIRSTYR / LASTYR	First (/last) data year that the code appeared in (Created by DataShare)
Y20YY	Code was present in specified year (0=No, 1=Yes)

E. INTERNATIONAL CLASSIFICATION OF DISEASES, 10TH REVISION, CLINICAL MODIFICATION (ICD-10-CM)

The International Classification of Diseases, Tenth Revision (ICD-10) system is published by the World Health Organization and includes diagnosis codes for diseases and related health problems. CMS and the National Center for Health Statistics maintain the “Clinical Modification” ICD-10-CM version.

ICD-10-CM codes are structured in a hierarchy, with additional digits after the decimal point signifying more specific codes. For example:

- A01 = “Typhoid and paratyphoid fevers”
- A01.0 = “Typhoid fever”
- A01.00 = “Typhoid fever, unspecified”
- A01.01 = “Typhoid meningitis”
- A01.02 = “Typhoid fever with heart involvement”

However, not all level of the hierarchy are considered official “billable” codes for CMS claims processing and other administrative systems. In the example above, “A01” and “A01.0” are not considered billable codes.

Table Name	ICD10CM
Source	UMLS - Metathesaurus (nih.gov)
Version currently in REFLIB	November 2022 (UMLS2022AB)
Release Frequency	Annual (published in July, first appears in Fall (AB) releases of the UMLS Metathesaurus)
Release-specific or cumulative?	Cumulative - DataShare combines multiple years of ICD-10-CM diagnosis code lists into one cumulative table in Oracle so that historical information and codes are retained.
Other notes or links	UMLS Metathesaurus - ICD10CM (International Classification of Diseases, Tenth Revision, Clinical Modification) - Synopsis (nih.gov) ICD- 10 - CM International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) (cdc.gov)

Variable name	Variable description
CODE	ICD-10-CM diagnosis code Note: In this table, the decimal point is included in the ICD diagnosis code. However, other administrative billing databases may not include the decimal, so users may need to remove the decimals before attempting to merge/join with this table.
DESCRIP	Text description of diagnosis for disease or health condition.
DESCRIP_ALT	Abbreviated text description of diagnosis for disease or health condition. Note: This field is missing for codes that are not considered “billable”. See note about ICD-10-CM hierarchy above.
EVER_ACTIVE	Was the code ever active during the period for which we have UMLS data? (Created by DataShare)
FIRST_RELEASE	UMLS version/release in which the code was first introduced (Created by DataShare)

F. INTERNATIONAL CLASSIFICATION OF DISEASES, 10TH REVISION, PROCEDURE CODING SYSTEM (ICD-10-PCS)

The International Classification of Diseases, Tenth Revision, Procedure Coding System (ICD-10-PCS) system is maintained by CMS and includes procedure codes for medical procedures, treatments, and services.

Table Name	ICD10PCS
Source	UMLS - Metathesaurus (nih.gov)
Version currently in REFLIB	November 2022 (UMLS2022AB)
Release Frequency	Annual (Fall (AB) releases of the UMLS Metathesaurus)
Release-specific or cumulative?	Cumulative - DataShare combines multiple years of ICD-10-PCS procedure code lists into one cumulative table in Oracle so that historical information and codes are retained.
Other notes or links	UMLS Metathesaurus - ICD10PCS (ICD-10 Procedure Coding System) - Synopsis (nih.gov) ICD-10 CMS

Variable name	Variable description
CODE	ICD-10-PCS procedure code Codes are 7-character alphanumeric. No decimals are used.
DESCRIP	Text description of procedure
DESCRIP_ALT	Abbreviated text description of procedure
EVER_ACTIVE	Was the code ever active during the period for which we have UMLS data? (Created by DataShare)
FIRST_RELEASE	UMLS version/release in which the code was first introduced (Created by DataShare)

III. OTHER DATA RESOURCES

DataShare also curates other reference data resources in REFLIB beyond the UMLS Metathesaurus. Most of these resources are publicly available, but for user convenience, we have uploaded and processed these files in PACE and made them available in REFLIB.

A. RESTRUCTURED BETOS CLASSIFICATION SYSTEM (RBCS)

The Restructured BETOS Classification System (RBCS) is a new classification standard based on the original BETOS standard (see above). Similar to BETOS, it aggregates HCPCS procedure codes into clinically meaningful categories of healthcare services.

REFLIB also has a **RBCS-BETOS crosswalk** file that links the original BETOS codes with the new RBCS codes and notes the timeframes when these matches apply.

Table Name	RBCS_YYYY
Source	Restructured BETOS Classification System - Centers for Medicare & Medicaid Services Data (cms.gov)
Version currently in REFLIB	2022
Release Frequency	Annual
Release-specific or cumulative?	Release-specific
Other notes or links	Restructured BETOS Classification System Data Dictionary - Centers for Medicare & Medicaid Services Data (cms.gov)

Variable name	Variable description
HCPCS_CD	HCPCS procedure code
RBCS_ID	RBCS identifier code
RBCS_CAT	Category code
RBCS_CAT_DESC	Category text description
RBCS_SUBCAT	Sub-category code
RBCS_SUBCAT_DESC	Sub-category text description
RBCS_FAMNUMB	RBCS Family Number
RBCS_FAMILY_DESC	RBCS Family text description
RBCS_MAJOR_IND	Major Procedure indicator (M=Major, O=Other, N=Non-procedure code)
HCPCS_CD_ADD_DT	HCPCS code effective date
HCPCS_CD_END_DT	HCPCS code end date
RBCS_ASSIGNMENT Eff_DT	Earliest data that RBCS ID code is effective.
RBCS_ASSIGNMENT End_DT	Latest date that the RBCS ID code can be applied.

B. ICD-10-CM AND ICD-10PCS GENERAL EQUIVALENCE MAPPINGS (GEMS)

The General Equivalence Mappings tables were created by CMS to aid in the transition from ICD9 to ICD10, which fully took effect in October 2015 in the Medicare system. These tables provide a crosswalk to translate ICD 9 diagnosis and procedure codes to the equivalent ICD-10-CM (diagnosis codes) and ICD-10-PCS (procedure codes), or vice-versa.

When using the GEMS table, it is suggested that researchers perform both forward mapping and backward mapping to confirm which codes to use from each system to capture the condition or procedure that they are trying to identify in claims data that span ICD9 and ICD10.

Table Name	ICD10CM_GEMS_{2010-2018}
Source	ICD-10 CMS
Version currently in REFLIB	2010 - 2018
Release Frequency	Retired. Annually from 2010-2018
Release-specific or cumulative?	Release-specific
Other notes or links	ICD-10 CMS 2018 ICD-10 CM and GEMs CMS (Other years also available here)

Variable name	Variable description
ICD9	ICD-9 diagnosis or procedure code
ICD10	ICD-10 diagnosis or procedure code
DIRECTION	Direction of the mapping - forwards (9>10) or backwards (10>9)
APPROXIMATE	0 = The complete meaning of the source and target codes are considered equivalent (i.e. a exact 1-to-1 match) in either direction 1 = The complete meaning of the source and target codes are <u>not</u> considered equivalent
NOMAP	0 = There is a mapping/translation for the code in the target file 1 = There is <u>no</u> mapping/translation for the code in the target file
COMBINATION	0 = Codes in the target file do not need to be combined with others to provide an equivalent to the source code 1 = A combination of codes in the target file is needed to provide an equivalent mapping to the source code
SCENARIO	In combination entries, enumerates a cluster of target codes that need to be combined to form an equivalent meaning of the source code
CHOICELIST	In combination entries, enumerates the list of one or more codes in the target system from which one code must be chosen to satisfy the equivalent meaning of the source code

C. CLINICAL CLASSIFICATION SOFTWARE - REFINED (CCSR) - DIAGNOSIS AND PROCEDURE VERSIONS

The Clinical Classification Software - Refined is a new classification scheme developed by the Healthcare Cost and Utilization Project (H-CUP) and the Agency for Healthcare Research and Quality (ARHQ). The CCSR facilitates aggregation of ICD-10-CM diagnosis codes or ICD-10-PCS procedure codes into clinically meaning groupings.

Tables for the original versions of Clinical Classification Software (CCS) are also available in REFLIB. More information available at [Research Tools \(ahrq.gov\)](https://www.ahrq.gov/research-tools) :

- CCS_CPT (CPT procedure codes)
- CCS_ICD10CM (ICD10 diagnosis codes)
- CCS_ICD10PCS (ICD10 procedure codes)
- CCS_ICD9CM_DX (ICD9 diagnosis codes, available in single-level and multi-level formats)
- CCS_ICD9CM_PX (ICD9 procedure codes, available in single-level and multi-level formats)

Table Name	CCSR_DX, CCSR_PX
Source	Clinical Classifications Software Refined (CCSR) (ahrq.gov)
Version currently in REFLIB	2022.1 (Released October 2021)
Release Frequency	Annual (October; valid for codes through September of the following year).
Release-specific or cumulative?	Release-specific
Other notes or links	Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses (ahrq.gov) Clinical Classifications Software Refined (CCSR) for ICD-10-PCS Procedures (ahrq.gov)

Variable name	Variable description
CCSR_DX:	
DX	ICD-10-CM Diagnosis Code
I10LABEL	ICD-10-CM Code Description
I10CCSRDEFTIP	Default CCSR Category Inpatient (IP)
I10CCSRRLABELDEFTIP	Default CCSR Category Description Inpatient (IP)
I10CCSRDEFTOP	Default CCSR Category Outpatient (OP)
I10CCSRRLABELDEFTOP	Default CCSR Category Description Outpatient (OP)
I10CCSR1	CCSR Category 1
I10CCSRRLABEL1	CCSR Category 1 Description
I10CCSR2	CCSR Category 2
I10CCSRRLABEL2	CCSR Category 2 Description
I10CCSR3	CCSR Category 3
I10CCSRRLABEL3	CCSR Category 3 Description
I10CCSR4	CCSR Category 4
I10CCSRRLABEL4	CCSR Category 4 Description
I10CCSR5	CCSR Category 5
I10CCSRRLABEL5	CCSR Category 5 Description
I10CCSR6	CCSR Category 6
I10CCSRRLABEL6	CCSR Category 6 Description
CCSR_PX:	

PX	ICD-10-PCS Procedure Code
I10LABEL	ICD-10-PCS Code Description
I10CCSR	PRCCSR
I10CCSRLABEL	PRCCSR Description
I10CCSRCD	Clinical Domain

D. CLINICAL CLASSIFICATION SOFTWARE – SERVICES AND PROCEDURES VERSION (CPT/HCPCS)

The Clinical Classification Software – Services and Procedures is classification scheme for CPT and HCPCS procedure codes developed by the Healthcare Cost and Utilization Project (H-CUP) and the Agency for Healthcare Research and Quality (AHRQ). The CCS-CPT facilitates aggregation of CPT and HCPCS procedure codes into clinically meaning groupings.

Table Name	CCS_CPT_CUMULATIVE
Source	HCUP-US Tools and Software Page CCS-Services and Procedures (ahrq.gov)
Version currently in REFLIB	2019.1 (includes codes from 2002-2019) through 2022.1 (Released May 2022)
Release Frequency	Annual (May; each version includes the codes that were valid during that calendar year).
Release-specific or cumulative?	Cumulative (Versions are merged together, with flags for each version that the code was present in)
Other notes or links	Users must accept the AMA CPT license agreement to download data and access documentation. https://hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccscpt_downloading.jsp CCS-SvcProc-User-Guide-v2022-1 (ahrq.gov)

Variable name	Variable description
CODE	CPT/HCPCS code
CCSPX	CCS procedure category
CCSPX_DESCRIP	CCS procedure category description
Y2019_v1	Flag indicating code was present in CCS CPT version 2019.1 (1=present, .=not present)
...	
Y2022_v1	Flag indicating code was present in CCS CPT version 2022.1 (1=present, .=not present)

E. PROVIDER OF SERVICE (POS)

CMS maintains the Provider of Service (POS) public use data file, which includes detailed information about hospital and non-hospital facilities, such as location, bed counts, and staffing. There are over 460 variables in this file, so users are encouraged to visit the links below for the data dictionary and other information about the POS files.

Table Name	POSYYYY
Source	Provider of Services Current Files CMS
Version currently in REFLIB	2011-2021
Release Frequency	Quarterly, but DataShare refreshes annually
Release-specific or cumulative?	Release-specific
Other notes or links	Provider of Services File - Hospital & Non-Hospital Facilities - Centers for Medicare & Medicaid Services Data (cms.gov)

Variable name	Variable description
	Please visit the “Resource” link on Provider of Services File - Hospital & Non-Hospital Facilities - Centers for Medicare & Medicaid Services Data (cms.gov) for the POS data dictionary.

F. MULTIPLE SEVERITY - DIAGNOSES RELATED GROUPS (MSDRG)

The Medicare-Severity Diagnosis-Related Group (MS-DRG) is a system used by CMS to classify hospital cases into groups of patients that are clinically similar and are expected to use the same level of hospital resources.

Table Name	MSDRG_YYYY
Source	MS-DRG Classifications and Software CMS
Version currently in REFLIB	2010-2020
Release Frequency	Annual
Release-specific or cumulative?	Release-specific
Other notes or links	FY 2022 IPPS Final Rule Home Page CMS (Table 5)

Variable name	Variable description
DRG	Diagnosis Related Group code
POST_ACUTE	Post-acute transfer care rule (Yes/No)
SPECIAL_PAY	Special payment methodology for post-acute care transfer (Yes/No)
MDC	Major Diagnosis Category
TYPE	Type (SURG or MED)
DESCRIP	Text description
WEIGHT	Weighting factor
LOS_GMEAN	Length of Stay - Geometric Mean
LOS_AMEAN	Length of Stay - Arithmetic Mean

G. NATIONAL PLAN AND PROVIDER ENUMERATION SYSTEM (NPPES)

The National Plan and Provider Enumeration system (NPPES) is maintained by CMS and provides information from the National Provider Identifier (NPI) program. NPI's are unique identifiers for providers and facilities. The NPPES data has 1 record per NPI and contains information about both active and deactivated NPIs, including names, other identifiers, geographic information, and physician specialty/taxonomy.

Table Name	NPPES_NPI
Source	NPI Files (cms.gov)
Version currently in REFLIB	October 2022
Release Frequency	Monthly but DataShare refreshes annually
Release-specific or cumulative?	Release-specific
Other notes or links	Data Dissemination CMS NPPES Data Dissemination - Readme.doc (cms.gov) National Plan and Provider Enumeration System (NPPES) NBER npi-to-sas/import_npi.sas · GitHub

VARIABLE NAME	Variable description
<i>NPPES has >300 fields, but many are repeated (e.g. OTHPID{1-50} = Other Provider Identifier 1-50)</i>	
See NPPES Data Dissemination - Readme.doc (cms.gov)	
CONDENSED LIST:	
NPI	National Provider Identifier (unique ID)
ENTITY	Entity type (1=Individual, 2=Organization)
REPLACEMENT_NPI	Replacement NPI
EIN	Employer Identification Number (EIN)
PORGNOME	Provider Organization Name (Legal Business Name)
PLNAME	Provider Last Name (Legal Name)
PFNAME	Provider First Name
PMNAME	Provider Middle Name
PNAMEPREFIX	Provider Name Prefix Text
PNAMESUFFIX	Provider Name Suffix Text
PCREDENTIAL	Provider Credential Text
PORGNOMEOTH	Provider Other Organization Name
PORGNOMEOTHCODE	Provider Other Organization Name Type Code
PLNAMEOTH	Provider Other Last Name
PFNAMEOTH	Provider Other First Name
PMNAMEOTH	Provider Other Middle Name
PNAMEPREFIXOTH	Provider Other Name Prefix Text
PNAMESUFFIXOTH	Provider Other Name Suffix Text
PCREDENTIALOTH	Provider Other Credential Text
PLNAMECODE	Provider Other Last Name Type Code
PMAILLINE1	Provider First Line Business Mailing Address
PMAILLINE2	Provider Second Line Business Mailing Address
PMAILCITYNAME	Provider Business Mailing Address City Name
PMAILSTATENAME	Provider Business Mailing Address State Name
PMAILZIP	Provider Business Mailing Address Postal Code

PMAILCOUNTRY	Provider Business Mailing Address Country Code (If outside U.S.)
PMAILTEL	Provider Business Mailing Address Telephone Number
PMAILFAX	Provider Business Mailing Address Fax Number
PLOCLINE1	Provider First Line Business Practice Location Address
PLOCLINE2	Provider Second Line Business Practice Location Address
PLOCCITYNAME	Provider Business Practice Location Address City Name
PLOCSTATENAME	Provider Business Practice Location Address State Name
PLOCZIP	Provider Business Practice Location Address Postal Code
PLOCCOUNTRY	Provider Business Practice Location Address Country Code (If outside U.S.)
PLOCTEL	Provider Business Practice Location Address Telephone Number
PLOCFAX	Provider Business Practice Location Address Fax Number
PENUMDATE	Provider Enumeration Date
LASTUPDATE	Last Update Date
NPIDEACTREASON	NPI Deactivation Reason Code
NPIDEACTDATE	NPI Deactivation Date
NPIREACTDATE	NPI Reactivation Date
PGENDER	Provider Gender Code
AOLNAME	Authorized Official Last Name
AOFNAME	Authorized Official First Name
AOMNAME	Authorized Official Middle Name
AOTITLE	Authorized Official Title or Position
AOTELNUM	Authorized Official Telephone Number
PTAXCODE{1-15}	Healthcare Provider Taxonomy Code 1-15
PLICNUM{1-15}	Provider License Number 1-15
PLICSTATE{1-15}	Provider License Number State Code 1-15
PPRIMTAX{1-15}	Healthcare Provider Primary Taxonomy Switch 1-15
OTHPID{1-50}	Other Provider Identifier 1-50
OTHPIDTY{1-50}	Other Provider Identifier Type Code 1-50
OTHPIDST{1-50}	Other Provider Identifier State 1-50
OTHPIDISS{1-50}	Other Provider Identifier Issuer 1-50
SOLEPROP	Is Sole Proprietor
ORGSUBPART	Is Organization Subpart
PARENT_ORG_LBN	Parent Organization Legal Business Name
PARENT_ORG_TIN	Parent Organization TIN
AONAME_PREFIX	Authorized Official Name Prefix Text
AONAME_SUFFIX	Authorized Official Name Suffix Text
AOCREDENTIAL	Authorized Official Credential Text
PTAXGROUP{1-15}	Healthcare Provider Taxonomy Group 1-15
CERTIFICATION_DATE	Certification Date

H. NATIONAL DRUG CODE DIRECTORY (NDC)

The National Drug Code Directory is produced by the FDA and indexes information about all drug products on the market *as of the release date of the dataset*.

Note: The NDC directory only includes information about finished drug products, unfinished drugs, and compounded drug products. Discontinued codes are not present in the NDC directory. For this reason, users are advised to consult other sources of drug code information if working with older data.

Table Name	NDC_PACKAGE, NDC_PRODUCT
Source	National Drug Code Directory FDA
Version currently in REFLIB	October 2022
Release Frequency	Monthly but DataShare refreshes annually
Release-specific or cumulative?	Release-specific
Other notes or links	NDC Product File Definitions FDA NDC Package File Definitions FDA Electronic Drug Registration and Listing System (eDRLS) FDA

VARIABLE NAME	Variable description
<u>NDC_PACKAGE table:</u>	
PRODUCTID	Product ID
PRODUCTNDC	Product NDC - Labeler code and product code segments of the NDC, separated by hyphens
NDCPACKAGECODE	NDC Package Code - Labeler code, product code, and package code segments, separated by hyphens
PACKAGEDESCRIPTION	Package description (size, type of packaging)
STARTMARKETINGDATE	Start date of product marketing
ENDMARKETINGDATE	End date of product marketing
NDC_EXCLUDE_FLAG	Indicator if package has been removed from NDC Directory (Values: E, U, I)
SAMPLE_PACKAGE	Indicator if package is distributed as a sample package (Y/N)
NDC	National Drug Code
<u>NCD_PRODUCT table:</u>	
PRODUCTID	Product ID
PRODUCTNDC	Product NDC - Labeler code and product code segments of the NDC, separated by hyphens
PRODUCTTYPENAME	Type of product (Cellular therapy, Human OTC drug, Human prescription drug, Non-standardized allergenic, Plasma derivative, Standardized allergenic, Vaccine)
PROPRIETARYNAME	Brand/Trade name
PROPRIETARYNAMESUFFIX	Suffix to trade name, such as "XR", "PM"
NONPROPRIETARYNAME	Generic name
DOSAGEFORMNAME	Dosage form
ROUTENAME	Route
STARTMARKETINGDATE	Start date of product marketing
ENDMARKETINGDATE	End date of product marketing
MARKETINGCATEGORYNAME	Marketing category

APPLICATIONNUMBER	Application number
LABELERNAME	Name of company
SUBSTANCENAME	Active ingredient list
ACTIVE_NUMERATOR_STRENGTH	Strength value for active ingredients
ACTIVE_INGRED_UNIT	Strength units for active ingredients
PHARM_CLASSES	Pharmacological class categories
DEASCHEDULE	DEA Schedule (CI, CII, CIII, CIV, CV)
NDC_EXCLUDE_FLAG	Indicator if package has been removed from NDC Directory (Values: E, U, I)
LISTING_CERTIFIED_THROUGH	Date when listing while expire if not updated or certified

I. AREA HEALTH RESOURCE FILES (ARHF)

Note: These files are very complex and do not have human-readable variable names (e.g. F20283). We list them here as a resource, but we do not currently have them in the REFLIB Oracle schema.

The Area Health Resource Files (AHRF) compiles a vast set of information about healthcare resources, such as workforce and facilities, aggregated at the county, state, or national level.

Table Name	AHRFYyyy
Source	Data Downloads (hrsa.gov)
Version currently in REFLIB	2019-2021
Release Frequency	Annual
Release-specific or cumulative?	Release-specific
Other notes or links	Area Health Resources Files (hrsa.gov)

VARIABLE NAME	Variable description
	Please refer to the Technical Documentation links on the Data Downloads page.

J. CROSSWALK FOR ZIP CODE TO HOSPITAL SERVICE AREA (HSA) AND HOSPITAL REFERRAL REGION (HRR)

The Dartmouth Atlas of Health Care project collects statistics at different levels of geographic aggregation. The ZIP-to-HRR file is a crosswalk that matches zip codes to a Hospital Service Area and Hospital Referral Region geographic area.

Table Name	HRRYYYY
Source	Supplemental Data - Dartmouth Atlas DATA
Version currently in REFLIB	1999-2019
Release Frequency	Annual
Release-specific or cumulative?	Release-specific
Other notes or links	Home - Dartmouth Atlas of Health Care Dartmouth Atlas DATA

VARIABLE NAME	Variable description
ZIP	5-digit zip code
HSA	Hospital service area
HRR	Hospital referral region

K. AREA DEPRIVATION INDEX (ADI) - 9-DIGIT ZIP CODE LEVEL

The Area Deprivation Index (ADI) is a ranking/measure of socioeconomic disadvantage aggregated at the census block group level. ADI is also reported at the 9-digit zip code level; REFLIB contains the zip code version.

Table Name	ADIYYYY
Source	https://www.neighborhoodatlas.medicine.wisc.edu/download
Version currently in REFLIB	2015 and 2019
Release Frequency	5-years
Release-specific or cumulative?	Release-specific
Other notes or links	Neighborhood Atlas - Home (wisc.edu)

VARIABLE NAME	Variable description
ZIP	9-digit ZIP code
ADI_NATLRANK	Area deprivation index - National ranking
ADI_STATERANK	Area deprivation index - State ranking
SSA_STATE	Social Security Administration (SSA) State code